Early-warning signals of topological collapse in interbank networks (Supplementary Information)

Tiziano Squartini
Instituut-Lorentz for Theoretical Physics, Leiden Institute of Physics, University of Leiden, Niels Bohrweg 2, 2333 CA Leiden, The Netherlands

Iman van Lelyveld
De Nederlandsche Bank, PO box 98, 1000 AB Amsterdam, The Netherlands

Diego Garlaschelli
Instituut-Lorentz for Theoretical Physics, Leiden Institute of Physics, University of Leiden, Niels Bohrweg 2, 2333 CA Leiden, The Netherlands
(Dated: October 9, 2013)

Supplementary Information for "Early-warning signals of topological collapse in interbank networks"

PACS numbers: Valid PACS appear here

I. DATA

Data are taken from the database used in [1] by using prudential reporting of balance sheet positions of Dutch banks. The data includes all the exposures between Dutch banks (from contractual obligations to swaps) up to one year and of more than 1.5 million euros, on a quarterly reported frequency (from 1998Q1 to 2008Q4) [1]. In other words, the data covers forty-four time-periods, corresponding to the forty-four ends of quarters of eleven years (from 1998 to 2008) and shows only the existence (or not) of exposures between (anonymized) Dutch banks of more than 1.5 million euros. Note that the last four temporal snapshots correspond to the year 2008, i.e. the fir st year of the self-evident crisis, whose beginning can traced back to August 2007 (as already pointed out in [1]), i.e. the sixth-last temporal snapshot considered here.

Given the nature of the available data, the Dutch interbank network (DIN) is represented as a *binary*, *directed network* where vertices represent banks and links represent exposures: a link pointing from bank i to bank j, at time t, indicates the existence of (at least) an exposure of more than 1.5 million euros, directed from i to j, registered at the end of the particular quarter t. The number N of banks varies from period to period, oscillating between 91 and 102 (see main text). For each quarter t (with $t = 1, \ldots, 44$), the structure of the network is therefore entirely described by an $N \times N$ (in general asymmetric) adjacency matrix A, whose entries are $a_{ij} = 1$ if a binary directed link from bank i to bank j exists during that quarter, and $a_{ij} = 0$ otherwise. The number L of directed links is therefore computed as

$$L \equiv \sum_{i=1}^{N} \sum_{j(\neq i)=1}^{N} a_{ij} \tag{1}$$

And the link density, or connectance c, is

$$c \equiv \frac{L}{N(N-1)} \tag{2}$$

II. THE CORE-PERIPHERY MODEL

In the literature about interbank markets (see e.g. ref. [1]), the Core-Periphery (CP) model is a popular axiomatic model, describing an ideal interbank network where nodes are perfectly split in two different classes: *core-nodes* and *periphery-nodes*. Thus, the core and the periphery are two non-overlapping sets of vertices, defined by means of the following three axioms [1]:

A1: core banks are all bilaterally linked with each other;

A2: periphery banks do not lend to each other;

A3: core banks both lend to and borrow from at least one periphery bank.

These axioms describe an ideal structure like that shown in fig. 1.

Of course, real interbank networks differ from a perfect core-periphery structure as postulated by the model. Nonetheless, it is still possible to look for the optimal partition of vertices between core and periphery, i.e. the partition which minimizes the number of deviations from the CP model. In particular, three types of deviations, or 'errors' can be defined, depending on which of the three axioms is violated [1]. For instance, if there is only one link between two banks that have been assigned to the core (instead of the postulated two links in the CP model), this is considered as one error of the first type. And if there is one link between two periphery banks, this is considered as one error of the second type. Finally, if a core bank does not lend (borrow) to the periphery at all this is assigned as many errors as there are periphery banks. For each of the three types of deviations, the 'error score' is simply the number of errors of that type. The total error score (for a given partition) is the sum of the scores of all three types of errors:

$$\epsilon \equiv \epsilon_1 + \epsilon_2 + \epsilon_3$$

The optimal partition is the one that minimizes ϵ, i.e. the one that best approximates the CP model.

We used a computational algorithm [1] that looks for optimal solutions. Given the resulting partition, we measured the number N_c of nodes in the core and the number $N_p = N - N_c$ of nodes in the periphery. Similarly, we measured the number of links in the core as

$$L_c \equiv \sum_{i \neq j,\ (i,j \in core)} a_{ij} \tag{3}$$

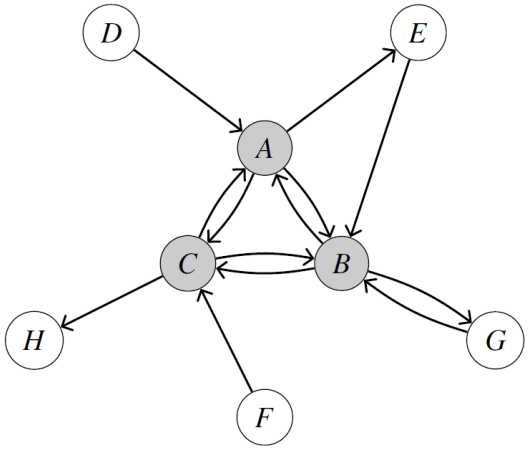

FIG. 1. An ideal example of core-periphery model: banks A, B, C belong to the core, and the other ones to the periphery. After ref. [1].

and the number of links in the periphery as

$$L_p = \sum_{i \neq j,\ (i,j \in periphery)} a_{ij} = L - L_c \tag{4}$$

(from now on, the symbol $\sum_{i \neq j}$ indicates the two nested sums $\sum_{i=1}^{N} \sum_{j(\neq i)=1}^{N}$). So we can distinguish between a *core-connectance* and a *periphery-connectance*, respectively defined as

$$c_c \equiv \frac{L_c}{N_c(N_c - 1)}, \quad c_p \equiv \frac{L_p}{N_p(N_p - 1)}. \tag{5}$$

The analysis of the above quantities is shown in the main text. In section VI we employ the CP model more intensively.

III. NULL MODELS

The approach presented here rests upon the *exponential random graphs* formalism [2–8]: given an appropriately chosen set of graphs (in what follows the *grandcanonical ensemble* $\mathcal{G}$, i.e. the collection of graphs with the same number of nodes of the observed network and a varying number of links, from zero to $N(N-1)$) a probability coefficient like the following

$$P(A|\vec{\theta}) = \frac{e^{-H(A,\,\vec{\theta})}}{Z(\vec{\theta})} \tag{6}$$

is associated with each of them. In eq. (6), $H(A,\,\vec{\theta}) \equiv \sum_a \theta_a \pi_a(A)$ is the *Hamiltonian* of the graph, i.e. the linear combination of topological contraints (dependent on the particular adjacency matrix, A) we choose to impose on the aforementioned ensemble [8] and the normalization constant, $Z(\vec{\theta}) \equiv \sum_{A \in \mathcal{G}} e^{-H(A,\,\vec{\theta})}$, is the *partition function*.

The unknown parameters can be estimated by maximizing the *log-likelihood* function of the network, $\ln \mathcal{L}(\vec{\theta}) = \ln P(A|\vec{\theta})$, with respect to the constraints [7, 8]

$$\left. \frac{\partial \ln \mathcal{L}(\vec{\theta})}{\partial \theta_a} \right|_{\vec{\theta}^*} \equiv 0,\ \forall\, a, \tag{7}$$

or, equivalently, by solving

$$\pi_a(A) = \langle \pi_a \rangle(\vec{\theta}^*) \equiv \langle \pi_a \rangle^*,\ \forall\, a \tag{8}$$

i.e. a list of equations imposing the values of the expected constraints to be equal to the observed ones (the term "expected", here, refers to the weighted average taken on $\mathcal{G}$, the weights being the probability coefficients above) [7, 8].

Once the numerical values $\vec{\theta}^*$ of the parameters have been determined, the expected value of any other topological quantity of interest, $X(A)$, is simply given by:

$$\langle X\rangle^* = \sum_{A\in\mathcal{G}} X(A)P(A|\vec{\theta}^*). \tag{9}$$

However, since the expected values of the most common quantities in complex networks theory are difficult to calculate exactly, it is often necessary to rely on the linear approximation method: $\langle X\rangle^* \simeq X(\langle A\rangle^*)$, with $\langle A\rangle^*$ indicating the expected adjacency matrix, whose elements will be indicated as $\langle a_{ij}\rangle^* \equiv p^*_{ij}$. By means of the same approximation, it is also possible to calculate the standard-deviation of the quantities of interest, as described in [8].

If the chosen contraints are linear in the adjacency matrix elements (i.e. of the form $\pi(A) = \sum_{i=1}^{N}\sum_{j(\neq i)=1}^{N} a_{ij}\theta_{ij}$) the expected entries become Bernoullian functions of the unknown parameters:

$$\langle a_{ij}\rangle = p_{ij} = \frac{e^{-\theta_{ij}}}{1+e^{-\theta_{ij}}} \equiv \frac{x_{ij}}{1+x_{ij}}. \tag{10}$$

The next two subsections will be devoted to the explanation of the null models considered for the present analysis.

A. Directed Random Graph Model

The *Directed Random Graph* (DRG in what follows) is the most well-known null model in complex networks theory. Its Hamiltonian is composed by only one addendum, the total number of links:

$$H(A,\,\vec{\theta}) = \alpha\, L = \sum_{i=1}^{N}\sum_{j(\neq i)=1}^{N} \alpha\, a_{ij}. \tag{11}$$

Being a linear constraint, the probability coefficients have the functional form shown in eq. (10)

$$p_{ij} \equiv p \equiv \frac{e^{-\alpha}}{1+e^{-\alpha}} \equiv \frac{x}{1+x} \tag{12}$$

whose unknown parameter can be estimated by solving the likelihood prescription

$$L(A) = \langle L\rangle^* = \sum_{i=1}^{N}\sum_{j(\neq i)=1}^{N} p^*_{ij} = \sum_{i=1}^{N}\sum_{j(\neq i)=1}^{N} \frac{x^*}{1+x^*}. \tag{13}$$

Eq. (13) can be immediately solved to give $p^* = \frac{L(A)}{N(N-1)} = c(A)$. So, the observed connectance is nothing more that the DRG probability that two any nodes be connected. The major drawback of this null model is that only one probability coefficient, p, accounts for the probability connections of every pair of nodes, thus ignoring their heterogeneity. A more refined model is presented in the next subsection.

B. Directed Configuration Model

The *Directed Configuration Model* (DCM in what follows) is characterized by the following Hamiltonian:

$$H(A,\,\vec{\theta}) = \sum_{i=1}^{N}\left(\alpha_i k_i^{out} + \beta_i k_i^{in}\right) = \sum_{i=1}^{N}\sum_{j(\neq i)=1}^{N} \left(\alpha_i + \beta_j\right) a_{ij} \tag{14}$$

(where $k_i^{in}(A) = \sum_{j(\neq i)=1}^{N} a_{ji}$ is the in-degree of node i, $k_i^{out}(A) = \sum_{j(\neq i)=1}^{N} a_{ij}$ is the out-degree of node i): it is again a linear function of the adjacency matrix elements, leading to probability coefficients of the form

$$p_{ij} \equiv \frac{e^{-\alpha_i-\beta_j}}{1+e^{-\alpha_i-\beta_j}} \equiv \frac{x_i y_j}{1+x_i y_j}. \tag{15}$$

The likelihood prescription for the DCM becomes [8]

$$\begin{cases} k_i^{in}(A) &= \langle k_i^{in}\rangle^* = \sum_{j(\neq i)} p_{ji}^* = \sum_{j(\neq i)} \frac{x_j^* y_i^*}{1+x_j^* y_i^*}, \ \forall\, i \\ k_i^{out}(A) &= \langle k_i^{out}\rangle^* = \sum_{j(\neq i)} p_{ij}^* = \sum_{j(\neq i)} \frac{x_i^* y_j^*}{1+x_i^* y_j^*}, \ \forall\, i \end{cases} \tag{16}$$

where the indices run from 1 to N. The in-degree and out-degree of a node are nothing more than the number of banks a vertex receives loans from and the number of banks a vertex lends to. Apart from considering trivial cases, the previous system can be solved only numerically.

C. Reciprocal Configuration Model

The third null model we considered is the *Reciprocal Configuration Model* (RCM in what follows), where each vertex has the same number of reciprocated, out-going non-reciprocated, and in-coming non-reciprocated links as in the observed network. In other words, the RCM incorporates not only the information about the number of (in- and outward) neighbors of a bank, but also the local reciprocity structure of each node, by means of three degree sequences defined by using the dyadic variables in eqs. (25-27) [8–10]:

$$\begin{cases} k_i^{\rightarrow}(A) &= \sum_{j(\neq i)} a_{ij}^{\rightarrow} \equiv \sum_{j(\neq i)} a_{ij}(1-a_{ji}), \ \forall\, i \\ k_i^{\leftarrow}(A) &= \sum_{j(\neq i)} a_{ij}^{\leftarrow} \equiv \sum_{j(\neq i)} a_{ji}(1-a_{ij}), \ \forall\, i \\ k_i^{\leftrightarrow}(A) &= \sum_{j(\neq i)} a_{ij}^{\leftrightarrow} \equiv \sum_{j(\neq i)} a_{ij}a_{ji}, \ \forall\, i \end{cases} \tag{17}$$

so that the resulting Hamiltonian becomes

$$H(A, \vec{\theta}) = \sum_{i=1}^{N} \left(\alpha_i k_i^{\rightarrow} + \beta_i k_i^{\leftarrow} + \gamma_i k_i^{\leftrightarrow}\right). \tag{18}$$

The first degree counts the number of links coming to node i and not having a reciprocal partner, the second degree counts the number of links going out from node i and not having a reciprocal partner and the third degree counts the number of links outgoing from, or ingoing to, node i and having a reciprocal partner. The RCM imposes the above sequences as contraints on the grandcanonical ensemble. As a consequence, the likelihood condition prescribes to solve the following system [8]:

$$\begin{cases} k_i^{\rightarrow}(A) &= \langle k_i^{\rightarrow}\rangle^* = \sum_{j(\neq i)} \frac{x_i^* y_j^*}{1+x_i^* y_j^* + x_j^* y_i^* + z_i^* z_j^*}, \ \forall\, i \\ k_i^{\leftarrow}(A) &= \langle k_i^{\leftarrow}\rangle^* = \sum_{j(\neq i)} \frac{x_j^* y_i^*}{1+x_i^* y_j^* + x_j^* y_i^* + z_i^* z_j^*}, \ \forall\, i \\ k_i^{\leftrightarrow}(A) &= \langle k_i^{\leftrightarrow}\rangle^* = \sum_{j(\neq i)} \frac{z_i^* z_j^*}{1+x_i^* y_j^* + x_j^* y_i^* + z_i^* z_j^*}, \ \forall\, i \end{cases} \tag{19}$$

Once the unknown parameters have been numerically determined, the probability coefficients $\langle a_{ij}^{\rightarrow}\rangle^* \equiv (p_{ij}^{\rightarrow})^*$, $\langle a_{ij}^{\leftarrow}\rangle^* \equiv (p_{ij}^{\leftarrow})^*$ and $\langle a_{ij}^{\leftrightarrow}\rangle^* \equiv (p_{ij}^{\leftrightarrow})^*$ can be used to calculate the expected value of all the topological properties of interest. Note that the usual degree sequences are preserved under the RCM, because

$$k_i^{out} = k_i^{\rightarrow} + k_i^{\leftrightarrow}, \ k_i^{in} = k_i^{\leftarrow} + k_i^{\leftrightarrow}. \tag{20}$$

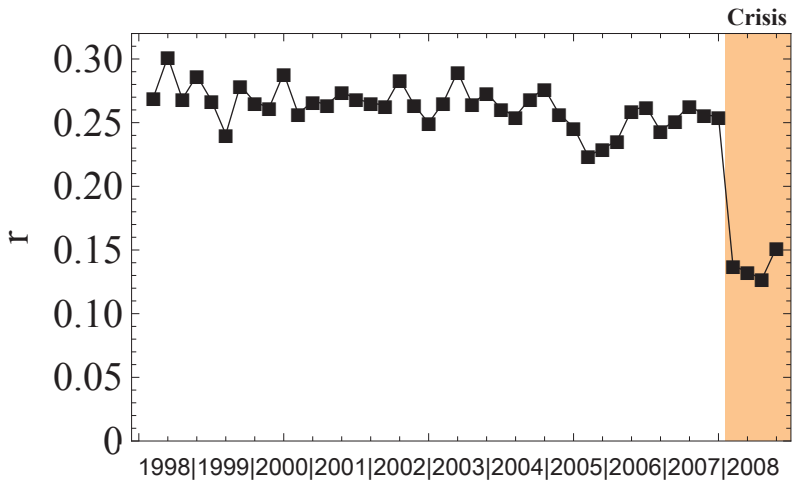

FIG. 2. Evolution of the reciprocity r of the interbank network.

IV. RECIPROCITY AND DYADS

The *reciprocity* is the fraction of links having a reciprocal partner (i.e. a link pointing in the opposite direction) [11], and is defined as

$$r \equiv \frac{\sum_{i=1}^{N}\sum_{j(\neq i)=1}^{N} a_{ij}a_{ji}}{L} \equiv \frac{L^{\leftrightarrow}}{L} \tag{21}$$

Unlike the number of vertices, the number of links, and the connectance, the reciprocity offers a clear signature of the crisis, as is evident in fig. 2. For most of the time period it shows an essentially constant trend, with small fluctuations around an average value of approximately 0.26, but the last four periods are characterized by an impressive decrease of the reciprocity value (approximately 40%): they lie almost 3 sigmas away from the sample average, clearly indicating that the DIN shows an anomalously low reciprocity value in those time periods already affected by the crisis.

What about the periods before it? Just looking at fig. 2, there is no strong evidence of the upcoming event, an thus the reciprocity by itself isn't sufficiently informative about the near future. However, the statistical significance of this conclusion can only be stated after a comparison with a well defined reference, i.e. the *null models* introduced in the previous section.

A. The ρ index

Fig. 2 shows the effectiveness of reciprocity r in characterizing the first year of the crisis. How statistically significant is the observed trend? To answer this question, let us implement the DRG and the DCM to compare the observed r with its expected value:

$$\langle r\rangle \equiv \frac{\langle L^{\leftrightarrow}\rangle}{\langle L\rangle} = \frac{\sum_{i\neq j} p_{ij}p_{ji}}{\sum_{i\neq j} p_{ij}} \tag{22}$$

In order to do this, let us calculate the ρ index [9], defined as

$$\rho \equiv \frac{r - \langle r\rangle}{1 - \langle r\rangle} \tag{23}$$

which automatically discounts for the effects of the imposed constraints. By definition, ρ ranges between 1 and -1: in fact, the denominator is always positive and, in magnitude, smaller than the numerator. It simply normalizes the

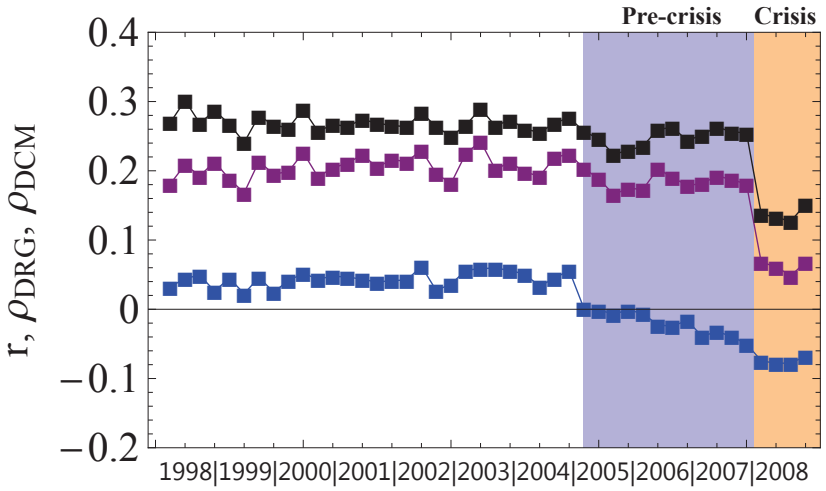

FIG. 3. Temporal evolution of the observed reciprocity r (black) and of ρ under the DRG (purple) and the DCM (blue).

index, not contributing to the sign of the quantity itself which, in turn, is decided only by the relative magnitude between the observed value r and its expectation. A positive sign indicates a stronger than expected tendency to reciprocate whereas a negative sign, a tendency weaker than expected to establish reciprocal links. The trends of ρ calculated under the DRG and the DCM are shown in fig. 3.

The positive sign of the trend of ρ under the DRG (i.e. ρ_{DRG}) indicates that the tendency of the network to reciprocate is stronger than expected under this model. This is intuitive by considering that $\langle r\rangle_{DRG} = p$ and that p is an average of all the single pair-specific probabilities, whose numerical value coincides with the observed connectance. In fact,

$$\langle r\rangle_{DRG} = \frac{\sum_{i\neq j} p^2}{\sum_{i\neq j} p} = p = \frac{\sum_{i\neq j} p_{ij}}{N(N-1)} = \frac{\langle L\rangle_{DRG}}{N(N-1)} = \\ = \langle c\rangle_{DRG} \tag{24}$$

and, by using the likelihood prescription, $\langle c\rangle_{DRG} = c(A)$ for any matrix A of the time period considered. Given the low value of the connectance, the DRG predicts a low value for r as well, such that $c < r$ for all the time-periods, thus underestimating the tendency of the links to establish mutual connections. Moreover, even if the DRG correctly describes the first year of the crisis (as evident by noting the final jump of ρ_{DRG}), this is exclusively due to the particular functional form of ρ_{DRG} itself, being a simple, small translation (and rescaling) of r towards lower values: $\rho_{DRG} = \frac{r-c}{1-c} \simeq r - c$. Thus, the discovery of patterns anticipating the crisis is trivially demanded to r which is, by itself, blind to this, as already pointed out. As a result, the network seems to suddenly depart from the initial (quite stable) configuration with many reciprocated links to the crisis configuration, where the number of reciprocal links sharply diminishes.

Far more interesting is the trend of ρ_{DCM}. As a general comment, the network is more consistent with the DCM null rather than with the DRG one, as the smaller values of the respective ρ indices show. In more detail, the DCM highlights two opposite patterns. During the first twenty-eight periods, the tendency of the network is to be reciprocated more than expected (similar to the DRG, but with the difference that ρ_{DCM} presents an almost constant trend, by showing smaller fluctuations than ρ_{DRG}): this implies that even the specification of the entire in- and out-degree sequences is not enough to fully account for the observed reciprocity, as the positive value of ρ_{DCM} witnesses. The same is valid also in the second sixteen periods, with the difference that the network inverts the tendency and tends to be less reciprocated than expected, showing an almost perfect monotonic decrease (with small, constant jumps in the value of ρ_{DCM}, of approximately four periods each). This clear anti-reciprocal behavior, not detected by the DRG but revealed by the DCM (i.e. not encoded in the total number of links, but partially encoded in the degree sequences) may be an early signature of the upcoming crisis, as the nodes start avoiding mutual exchanges two years before the 2008. However, the absence of significance bounds for ρ prevents us from drawing definitive conclusions

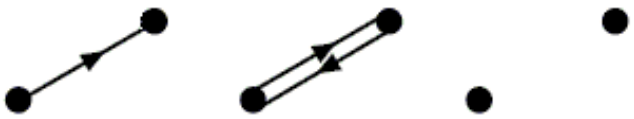

FIG. 4. The 3 possible binary, directed dyads.

about this point. Moreover, ρ is a generic index, only describing the global tendency of links to reciprocate or not. This leads to the question how the single pairs of nodes behave, as discussed in the next section.

B. Dyadic motifs

In order to address the above points, we carried out a more detailed analysis of the reciprocity structure of the DIN, by looking at the possible *dyadic motifs*, i.e. the three ways any two given nodes can be (dis)connected in a binary, directed network (see fig. 4). Dyads provide detailed information about the local reciprocity structure of a network, by measuring how many single, or mutual, connections a given node has. The statistical significance of the ρ index temporal evolution can be confirmed by checking the statistical significance of the single dyads' behavior, which can in turn be measured by means of z-scores [8, 10]. The number $L^{\rightarrow}$ of non-reciprocated (single) dyads, (twice) the number $L^{\leftrightarrow}$ of reciprocated (full) dyads, and (twice) the number $L^{\nrightarrow}$ of empty dyads are defined below, along with the corresponding z-scores:

$$L^{\rightarrow} \equiv \sum_{i \neq j} a_{ij}(1 - a_{ji});\ z_{L^{\rightarrow}} = \frac{L^{\rightarrow} - \langle L^{\rightarrow} \rangle}{\sigma[L^{\rightarrow}]}, \tag{25}$$

$$L^{\leftrightarrow} \equiv \sum_{i \neq j} a_{ij} a_{ji};\ z_{L^{\leftrightarrow}} = \frac{L^{\leftrightarrow} - \langle L^{\leftrightarrow} \rangle}{\sigma[L^{\leftrightarrow}]}, \tag{26}$$

$$L^{\nrightarrow} \equiv \sum_{i \neq j} (1 - a_{ij})(1 - a_{ji});\ z_{L^{\nrightarrow}} = \frac{L^{\nrightarrow} - \langle L^{\nrightarrow} \rangle}{\sigma[L^{\nrightarrow}]}. \tag{27}$$

The expected value and sigma of the dyads can be calculated analytically under both the DRG and the DCM considering that, for networks with local constraints, the dyads in a network are independent random variables [8]. The temporal evolution of the dyadic z-scores under the DRG and the DCM is portrayed in the Main Text (note that, under the RCM, each of the three dyadic abundances is reproduced by construction, which trivially implies that all dyadic z-scores are zero).

The dyadic z-scores show the same behaviour as the ρ index. The explanation lies in the analytical form of such quantites: the numerators of the ρ index and the dyads are (except for a minus sign) the same. In fact

$$\rho = \frac{r - \langle r \rangle}{1 - \langle r \rangle} = \frac{\frac{L^{\leftrightarrow}}{L} - \frac{\langle L^{\leftrightarrow} \rangle}{L}}{1 - \frac{\langle L^{\leftrightarrow} \rangle}{L}} = \frac{L^{\leftrightarrow} - \langle L^{\leftrightarrow} \rangle}{L - \langle L^{\leftrightarrow} \rangle} \tag{28}$$

and, considering that $L = L^{\leftrightarrow} + L^{\rightarrow}$,

$$L^{\rightarrow} - \langle L^{\rightarrow} \rangle = -(L^{\nrightarrow} - \langle L^{\nrightarrow} \rangle) = -(L^{\leftrightarrow} - \langle L^{\leftrightarrow} \rangle). \tag{29}$$

Note that, under the RCM, all local and global dyadic properties of the real network are preserved. As a consequence, $z_{L^{\leftrightarrow}}$, $z_{L^{\rightarrow}}$ and $z_{L^{\nrightarrow}}$ are, by construction, zero. For the same reason, we also have $\rho_{RCM} = 0$, and the observed reciprocity structure is completely reproduced by the model. So, the RCM fixes the dyadic motifs and can reveal patterns of self-organization *between* dyads, pointing out how triads (or more numerous sets) of nodes interact.

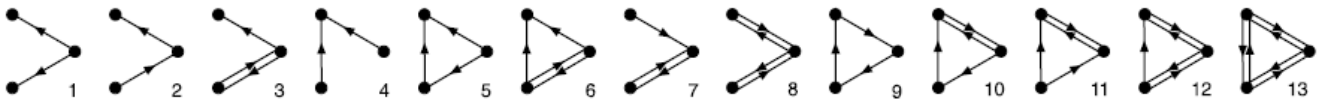

FIG. 5. The 13 possible triadic motifs involving three connected vertices.

V. TRIADS

The 13 possible motifs involving three connected vertices are shown in fig. 5. The analysis of these triadic motifs has been carried out by implementing the DCM and the RCM, as shown in the Main Text. Here we provide details of that analysis, and more results relating to the DRG.

A. Triadic z-scores

For each motif $m = 1, \dots 13$, the abundance N_m (up to a constant factor α_m that depends on the symmetry of the particular motif, which will drop out of all measured quantities) is obtained as shown in table I.

Triadic motif (m)	Abundance (N_m)
1	$\sum_{i\neq j\neq k}(1-a_{ij})a_{ji}a_{jk}(1-a_{kj})(1-a_{ik})(1-a_{ki})$
2	$\sum_{i\neq j\neq k}a_{ij}(1-a_{ji})a_{jk}(1-a_{kj})(1-a_{ik})(1-a_{ki})$
3	$\sum_{i\neq j\neq k}a_{ij}a_{ji}a_{jk}(1-a_{kj})(1-a_{ik})(1-a_{ki})$
4	$\sum_{i\neq j\neq k}(1-a_{ij})(1-a_{ji})a_{jk}(1-a_{kj})a_{ik}(1-a_{ki})$
5	$\sum_{i\neq j\neq k}(1-a_{ij})a_{ji}a_{jk}(1-a_{kj})a_{ik}(1-a_{ki})$
6	$\sum_{i\neq j\neq k}a_{ij}a_{ji}a_{jk}(1-a_{kj})a_{ik}(1-a_{ki})$
7	$\sum_{i\neq j\neq k}a_{ij}a_{ji}(1-a_{jk})a_{kj}(1-a_{ik})(1-a_{ki})$
8	$\sum_{i\neq j\neq k}a_{ij}a_{ji}a_{jk}a_{kj}(1-a_{ik})(1-a_{ki})$
9	$\sum_{i\neq j\neq k}(1-a_{ij})a_{ji}(1-a_{jk})a_{kj}a_{ik}(1-a_{ki})$
10	$\sum_{i\neq j\neq k}(1-a_{ij})a_{ji}a_{jk}a_{kj}a_{ik}(1-a_{ki})$
11	$\sum_{i\neq j\neq k}a_{ij}(1-a_{ji})a_{jk}a_{kj}a_{ik}(1-a_{ki})$
12	$\sum_{i\neq j\neq k}a_{ij}a_{ji}a_{jk}a_{kj}a_{ik}(1-a_{ki})$
13	$\sum_{i\neq j\neq k}a_{ij}a_{ji}a_{jk}a_{kj}a_{ik}a_{ki}$

TABLE I. Classification and abundances (up to a symmetry factor) of the 13 triadic motifs. The three nested sums run from 1 to N.

The z-score for the abundance of a particular triadic motif reads

$$z_m = \frac{\alpha_m N_m - \langle \alpha_m N_m \rangle}{\sigma[\alpha_m N_m]} = \frac{N_m - \langle N_m \rangle}{\sigma[N_m]}, \tag{30}$$

B. Triadic structure under the DRG

Under the DRG, the z-scores are easy to calculate analytically, considering that $p_{ij} = p$, $\forall\, i \neq j$ and by using the linear approximation to calculate the standard-deviations (as shown in ref. [8]):

$$\langle N_m \rangle = T_1 p^k (1-p)^{6-k},\ \sigma_{N_m} = T_2 \left[k p^{k-1}(1-p)^{6-k} - (6-k)p^k(1-p)^{5-k} \right] \tag{31}$$

where $k = 2, 3, 4, 5$ respectively for motif 2, 5 and 9, 10, 12, $T_1 \equiv N(N-1)(N-2)$ is the number of distinct, directed triads and $T_2 \equiv (N-2)\sqrt{N(N-1)p(1-p)}$. The motifs considered here are the ones already considered in the Main Text: motifs 2, 5, 9, 10 and 12. The results are shown in figs. 6, 7 and 8.

The z-scores under the DRG do not single out any further division in sub-periods, as shown in fig. 6: motif 12 is the only one that displays variations, but the latter correspond to a non-monotonic, oscillating temporal trend, as shown in fig. 7. Motifs 2, 10 and 12 do not even signal the anomalous period of the crisis, as evident by looking

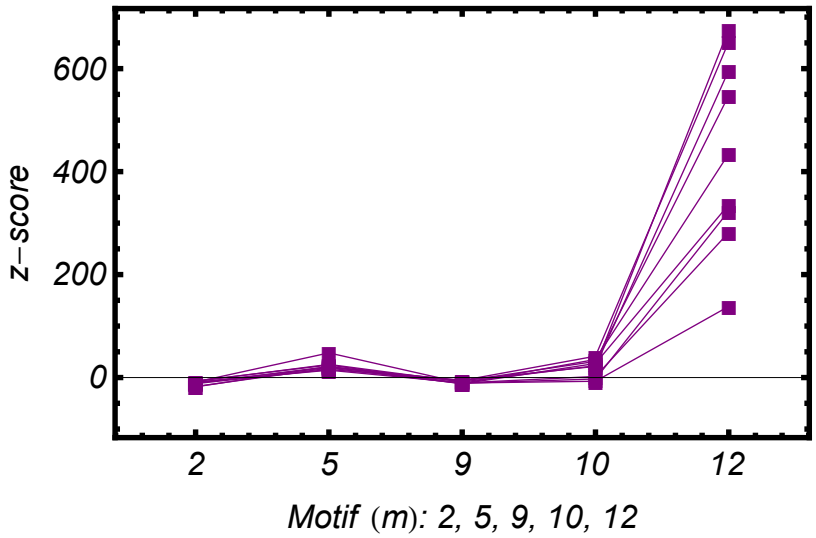

FIG. 6. Triadic z-scores for the motifs 2, 5, 9, 10, 12 in the quarterly snapshots 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, under the DRG.

at fig. 7: in fact, no significant departure from the previous periods-trend is appreciable. In this respect, the only useful trend is provided by motif 5, which shows an evident increase just before the beginning of the critical period. However, no evidence of a pre-crisis period is detectable in any of the four considered motifs, confirming the limited use of the DRG in providing useful predictions. So the above triadic motifs calculated under the DRG can at best confirm the same unpredictability as the dyadic ones (some triadic patterns undergo a sudden change in 2008), and at worst show fluctuations which are so large during the entire period, that it is difficult to tell whether any significant change is taking place at all at the onset of the crisis.

Only motif 9 shows a more informative temporal evolution and partly confirms (even if with much less significance) the results we found for the same motif under the DCM and the RCM (see main text), as fig. 8 shows. Even if neither the crisis, nor the pre-crisis period are detected, the anomaly in the 'cyclic anomaly' period is still visible, appearing as a global increase of the z-scores' values before coming back to an almost constant value in the last temporal periods $t_{30} - t_{44}$.

C. Triadic structure under the DCM

Motifs' profiles under the DCM represent one of the most significant results of the whole analysis, unambiguously showing the evolution of the DIN, as clear by looking at the forty-four motifs' profiles plotted together in fig. 9. The apparent disorder resulting from plotting all the forty-four profiles together hides the four, stationary sub-profiles shown in the main text. We will discuss each of the sub-periods in turn.

In the first subperiod (covering the time-periods between t_1 and t_{10} - see main text), the DCM seems to explain quite well the motifs' profiles (almost all the z-scores lie between $z = +3$ and $z = -3$): the only outlier is motif 8, whose abundance is underestimated by the DCM (the z-score is positive). Looking back at fig. 5 this means that, even if the total number of reciprocal and empty dyads is consistent with the DCM prediction in this subperiod, the abundance of triads composed by two reciprocated dyads and one empty dyad is still underestimated by it. This suggests that the simple topological information about the number of neighbors is still not sufficient to account for this kind of dyadic interactions and more precise information about the nodes' local reciprocity structure is needed. In the second subperiod (between t_{11} and t_{18}), even if motif 8 is again underestimated, the biggest, evident discrepancy is between the abundance of motif 9 (rising with respect to the first ten years) and its expected value, while the other motifs remain quite stable. Motif 9 is a complete triad of single dyads. Its underestimation could be due to the presence of genuine self-organization patterns at the triadic level rather than the reciprocity structure being ignored. As in the previous case, the DCM does not account for them. The third subperiod (between t_{19} and t_{40}) shows several motifs becoming more significant than in the previous subperiods: motif 2 and motif 5 become overrepresented (i.e. underestimated by the DCM) and motifs 10 and 12 become underrepresented (i.e. overestimated by the DCM). In the fourth subperiod (between t_{41} and t_{44}), these patterns become even more pronounced. The increasing divergence of these motifs from the DCM null could be another signature of the crisis, also by direct comparison with the dyadic

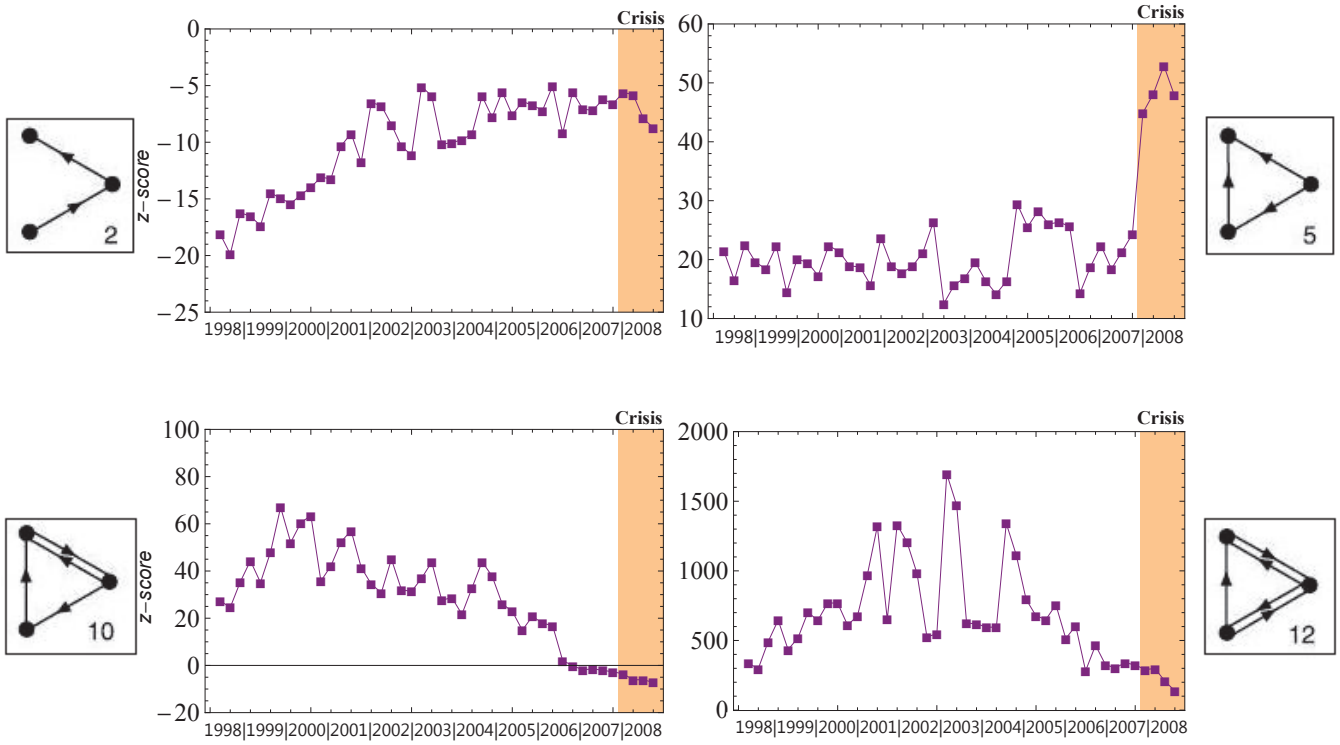

FIG. 7. Temporal evolution of motif 2 (top-left), motif 5 (top-right), motif 10 (bottom-left), motif 12 (bottom-right), under the DRG.

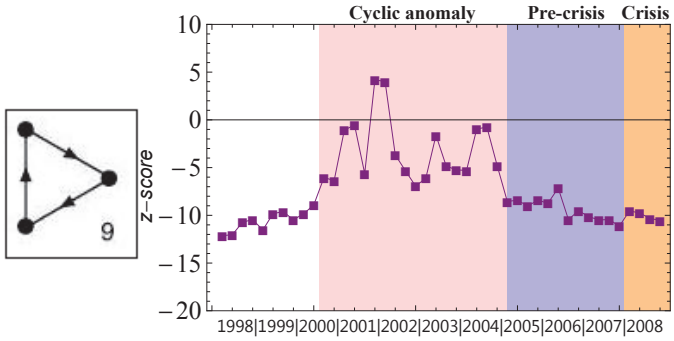

FIG. 8. Temporal evolution of motif 9, under the DRG.

abundance: as the number of reciprocal dyads is overestimated in the year 2008, so motif 10 and 12 are, both defined by (at least) one reciprocal dyad; as the number of single dyads is underestimated in the year 2008 (becoming only marginally consistent with the DCM prediction) so motif 2 and 5 are, both defined by (at least) two single dyads.

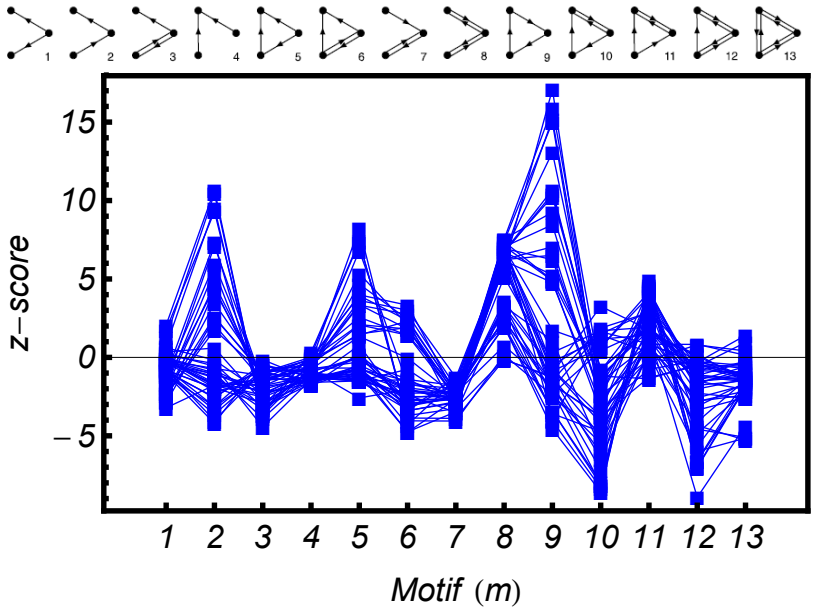

FIG. 9. Triadic z-scores for all 44 quarters, under the DCM.

D. Triadic structure under the RCM

What clearly emerges from the previous analysis is the fundamental role of the reciprocity structure of the DIN in describing both the crisis and the pre-crisis period. The DCM only partially accounts for the dyadic and the triadic structure of the network, highlighting the emergence of patterns not encoded in the degree sequences. In order to discover the presence of higher-order patterns, not encoded in the dyadic structure, the next step is to fix this kind of topological information beforehand as well. In other words, to disentangle dyadic effects from a genuine organization at the triadic level, we need to wash out the information about the reciprocity itself, by including it in our null model from the start. This results in the RCM with figure 10 showing the associated triadic z-scores. It is clear at a glance that now, except motifs 9 and 10, all motifs are approximately consistent with the null model. This means that, after controlling for the dyadic patterns already identified, motifs 9 and 10 still emerge as strongly significant triadic building blocks, irreducible to a combination of dyads.

The temporal evolution of these triadic z-scores (see main text) reveals that, out of the four motifs that are significant under the DCM in the third ('pre-crisis' phase) subperiod. Only motif 10 retains the same significance under the RCM, signalling that the information about the reciprocity is not sufficient to explain the abundance of this profile. Just as for motif 9, which sharply inverts its trend, motif 10 is composed of two single dyads and one reciprocal dyad, closing a triangle loop: this seems to confirm the additional presence of non-trivial third-order correlations. In other words, the network evolves from configurations with an exceptional, unexplained abundance of unreciprocated triangular loops to configurations where nodes strongly prefer avoiding them. The fourth ('crisis') subperiod is, again, a further evolution of the third one, showing motifs 9 and 10 evolving towards more strongly significant values. So, the first year of the crisis seems to be characterized by the strong absence of triangle interactions, while all the other patterns seem to be compatible with the RCM prediction. The evolution of motif 9 (see main text) confirms this, clearly indicating that a sort of triadic self-organization indeed exists and plays a fundamental role in shaping the interbank exchanges.

VI. EVOLUTION OF THE CORE-PERIPHERY STRUCTURE

In this section we describe our analyses of the evolution of the core-periphery structure of the DIN. As described in section II, we first looked for the the optimal partition of vertices providing the closest approximation to the CP model. As for the dyadic and triadic structure, we are interested in understanding whether the observed core-periphery structure is statistically significant, or whether it can be explained merely in terms of the local topological properties.

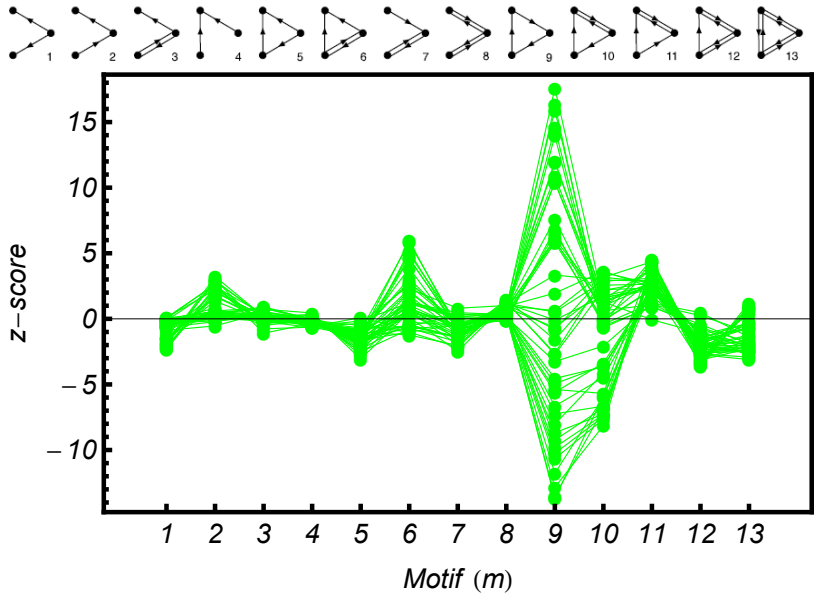

FIG. 10. Triadic z-scores for all 44 quarters, under the RCM.

A. Error score

To this end, we first studied whether the error score ϵ measured on the real network (given the optimal partition) is consistent with the value $\langle\epsilon\rangle$ (given the same partition) expected under the null models considered so far. Given a null model, we therefore define the z-score

$$z_\epsilon = \frac{\epsilon - \langle\epsilon\rangle}{\sigma[\epsilon]} \tag{32}$$

measuring by how many standard deviations the observed divergence from the ideal CP model differs from its expectation. As we now show, it is possible to evaluate z_ϵ analytically.

Generally, the optimal partition does not contain errors of the third type, since the latter are severely punished (See section II for the definition). As a consequence, it is sensible and easier to restrict our attention to the errors of the first and second type, which measure the deviation from a 'relaxed' CP model without the third axiom. If $V_c \equiv N_c(N_c-1)$ and $V_p \equiv N_p(N_p-1)$ denote the *volume* (number of possible links) of the core and periphery respectively, we consider the following simplified error score:

$$\epsilon \equiv \epsilon_1 + \epsilon_2 \equiv V_c - L_c + L_p \tag{33}$$

where $V_c - L_c$ is the number of missing (with respect to the CP model) links in the core and L_p is the number of extra (with respect to the CP model) links in the periphery. In an ideal core-periphery model we would have $L_c = V_c$ and $L_p = 0$, so that $\epsilon = 0$. Using the expression above, we can write

$$\langle\epsilon\rangle = V_c - \langle L_c\rangle + \langle L_p\rangle \tag{34}$$

Moreover, it is easy to check that

$$\sigma^2[\epsilon] = \sigma^2[L_c] + \sigma^2[L_p] \tag{35}$$

Taken together, these expression yield the desired analytical formula

$$z_\epsilon = \frac{\epsilon - \langle\epsilon\rangle}{\sigma[\epsilon]} = \frac{(L_p - L_c) - \langle L_p - L_c\rangle}{\sqrt{\sigma^2[L_c] + \sigma^2[L_p]}}, \tag{36}$$

Equation (36) shows that the fundamental variable appearing in the error definition is the difference between the links in the core and those in the periphery.

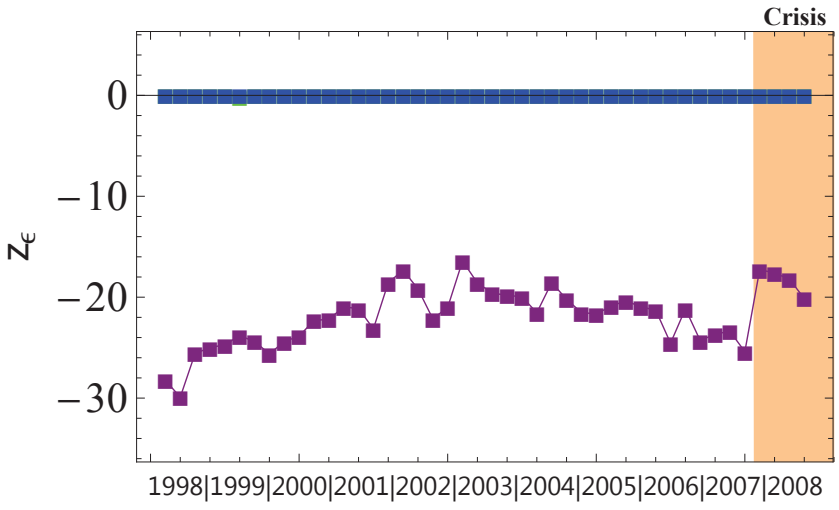

FIG. 11. z-score for the error score ϵ under the DRG (purple), the DCM (blue) and the RCM (green).

Armed with the above result, we employ the DRG, the DCM and the RCM and evaluate the corresponding z-scores. The results are shown in fig. 11. Using the DRG, we find that the observed core-periphery structure is strongly significant, as the DRG predicts a much larger expected error score resulting in larger negative values of z_ϵ. By contrast, using the DCM and RCM (which both give zero z_ϵ), we find that the real network is as close to the ideal CP model as are null models. The reason for z_ϵ being almost identically zero is that, as can be easily checked, all the possible rewiring moves that preserve the degrees of vertices *exactly* also preserve the number of errors of the first and second type (note that this would imply $\langle\epsilon\rangle = 0$ and $\sigma[\epsilon]$, making z_ϵ undefined). Our different approach, where degrees are preserved *on average* (see section III), allows for errors not to be preserved exactly in each single realization of the network. This implies that $\sigma[\epsilon] > 0$, making z_ϵ properly defined. Still, we find that $z_\epsilon \approx 0$, meaning that $\langle\epsilon\rangle$ is extremely close to zero. In other words, the observed core-periphery structure, as measured from the error score, is not a genuinely higher-order property, since it is simply explained by the degrees of vertices (for a discussion of this result see also [1]).

B. Density contrast

To double-check the above results, we introduced an alternative measure of the strength of core-periphery structure, namely the *density contrast* defined as

$$\Delta c \equiv c_c - c_p = \frac{L_c}{N_c(N_c-1)} - \frac{L_p}{N_p(N_p-1)} = \frac{L_c}{V_c} - \frac{L_p}{V_p} \tag{37}$$

i.e., the difference between the link density in the core and the link density in the periphery (see section II). This quantity is a very intuitive measure of the excess core density and, unlike the error score, is not trivially preserved by rewiring moves that preserve degrees, either exactly or on average. The z-score for the density contrast is

$$z_{\Delta c} = \frac{\Delta c - \langle\Delta c\rangle}{\sigma[\Delta c]} = \frac{\left(\frac{L_c}{V_c} - \frac{L_p}{V_p}\right) - \left(\frac{\langle L_c\rangle}{V_c} - \frac{\langle L_p\rangle}{V_p}\right)}{\sqrt{\frac{\sigma^2_{L_c}}{V_c^2} + \frac{\sigma^2_{L_p}}{V_p^2}}}. \tag{38}$$

The evolution of $z_{\Delta c}$ under the three null models is shown in fig. 12. Clearly, the DRG prediction underestimates the contrast. This is easily understandable by looking back at fig. 1 in the Main Text: the only information the DRG uses is the global connectance, thus predicting a more homogeneous structure and a smaller difference between the two zones' density than observed. However, the first year of the crisis shows a jump towards smaller values of the z-score as if, exactly as pointed out by the reciprocity index ρ, in 2008 the network became sparser and actually

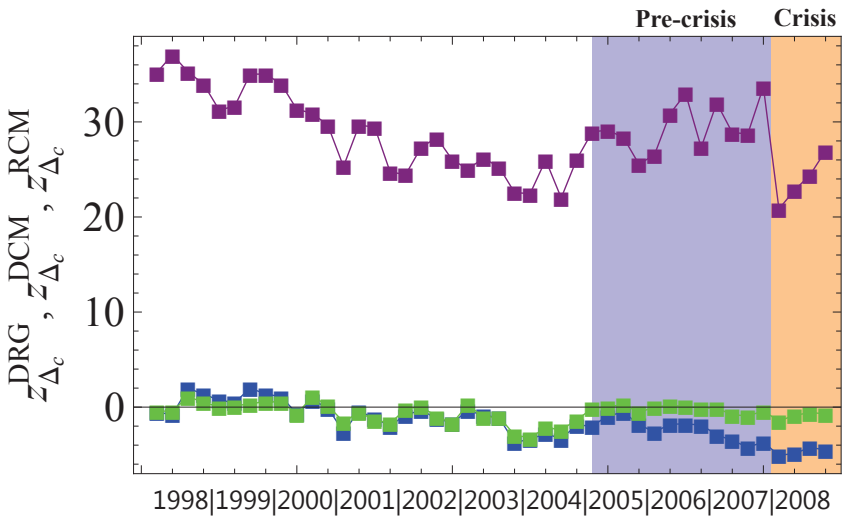

FIG. 12. z-score for the density contrast Δc under the DRG (purple), the DCM (blue) and the RCM (green).

more homogeneous (less difference between a "suspected" core and periphery), thus being in better agreement with the DRG.

The DCM and the RCM predict the same density contrast for most of the time period. The two models start deviating from each other approximately around period t_{30}, when the "pre-crisis" period starts: in particular, the DCM does not correctly account for the observed Δc, especially in the last periods (approximately two years, from t_{37} to t_{44}), when the crisis spreads. Here, the trend of the z-score highlights that the DIN structure is actually more homogeneous than predicted by the DCM: the network, as it evolves towards the crisis, seems to actually lose a sort of internal structure, based on the different density of links in the core and periphery areas. Moreover, this information seems to be again encoded into the reciprocity structure of the network as the the RCM, which accounts for this, correctly explains the density contrast. This sheds new light on the evolution of the reciprocity itself, as pointed out in the conclusions.

C. The distribution of $L_p - L_c$

The z-scores have a well defined statistical significance levels only in the case of normally-distributed variables. We tested this assumption only for a single time period (t_8), by proceeding in the following way. We implemented the DCM numerically for the chosen time period by generating 50.000 binary, directed matrices, according to the DCM rule

$$\begin{cases} a_{ij} = 1, & \text{if } u \sim U[0,1] \leq p^*_{ij} = \frac{x^*_i y^*_j}{1+x^*_i y^*_j} \\ a_{ij} = 0, & \text{else} \end{cases}$$

i.e. by extracting a real number uniformly distributed between 0 and 1 and comparing it with p^*_{ij} (whose numerical value was computed according to the maximum of the likelihood procedure), for each entry of the adjacency matrix, a_{ij}. Then, for each matrix M belonging to this numerically-generated grandcanonical ensemble, $\mathcal{M}$, we also calculated the distribution of the random variable $L_p - L_c$, its arithmetic mean,

$$\overline{L_p - L_c} = \frac{\sum_{M \in \mathcal{M}} L_p(M) - L_c(M)}{|\mathcal{M}|} \tag{39}$$

and its standard deviation,

$$\sigma^2[L_p - L_c] = \overline{(L_p - L_c)^2} - \left(\overline{L_p - L_c}\right)^2 . \tag{40}$$

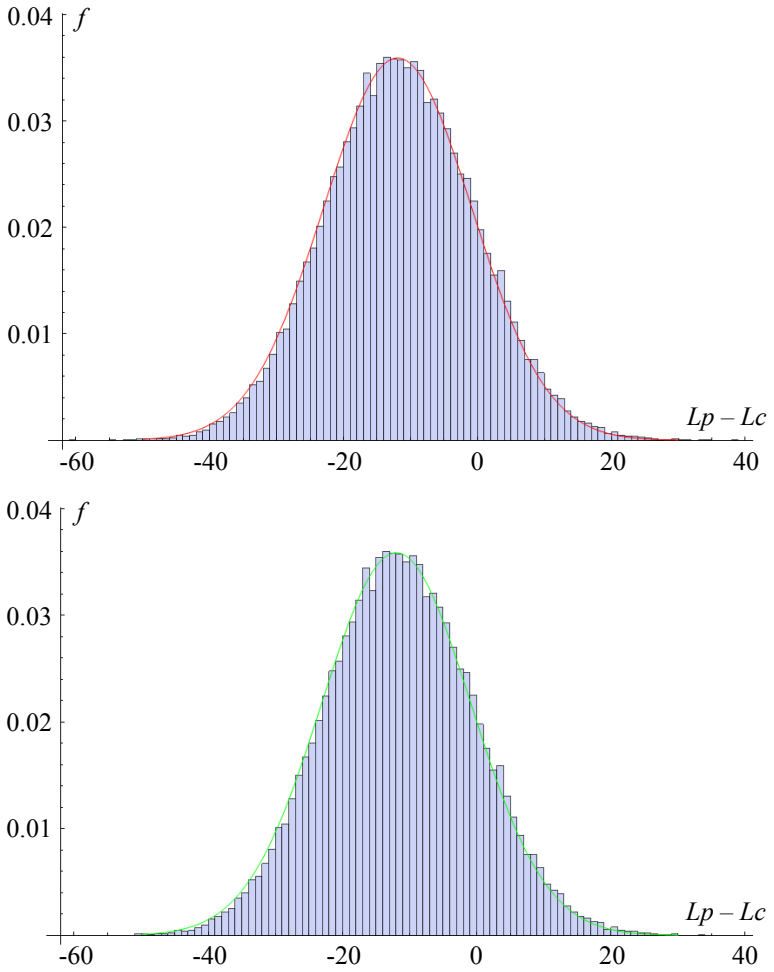

FIG. 13. Histogram showing the distribution of $L_p - L_c$ on $\mathcal{M}$ (a numerically-generated grandcanonical ensemble according to the DCM probability coefficients and composed by 50.000 matrices), Gaussian fit with parameters $\mu = \overline{L_p - L_c}$ and $\sigma^2 = \overline{(L_p - L_c)^2} - \left(\overline{L_p - L_c}\right)^2$ (top, red) and Gaussian fit with parameters $\mu = \langle L_p - L_c \rangle$ and $\sigma^2 = \langle (L_p - L_c)^2 \rangle - \langle L_p - L_c \rangle^2$ (bottom, green).

The results are shown in fig. 13. The histograms both show the same distribution of $L_p - L_c$: we simply added two fits. The red fit is a Gaussian fit with parameters $\mu = \overline{L_p - L_c}$ and $\sigma^2 = \overline{(L_p - L_c)^2} - \left(\overline{L_p - L_c}\right)^2$, i.e. the same as in eqs. (39) and (40). The green fit is again a Gaussian, with parameters $\mu = \langle L_p - L_c \rangle$ and $\sigma^2 = \langle (L_p - L_c)^2 \rangle - \langle L_p - L_c \rangle^2$, i.e. those analytically calculated by means of the maximum of the likelihood procedure (and implemented for the time-period t_8). What we observe is a substantial agreement between the distribution of $L_p - L_c$ calculated on the numerically-generated grandcanonical ensemble and the two Gaussian fits, thus confirming the usual statistical interpretation of the z-scores.

VII. AKAIKE'S INFORMATION CRITERION (AIC)

For the analysis of the DIN we have used three null models: the DRG, the DCM and the RCM. Which of these is the 'best' null model in terms of an optimal trade-off between parsimony and ability to replicate the data? In order to answer this question, we use the *Akaike's Information Criterion* (AIC in what follows) and the *Akaike weights* [12]. These techniques have been recently used in the analysis of other networks [13].

For each snaphot t and for each model, AIC is defined as the difference between (twice) the number K of parameters and the log-likelihood at its maximum:

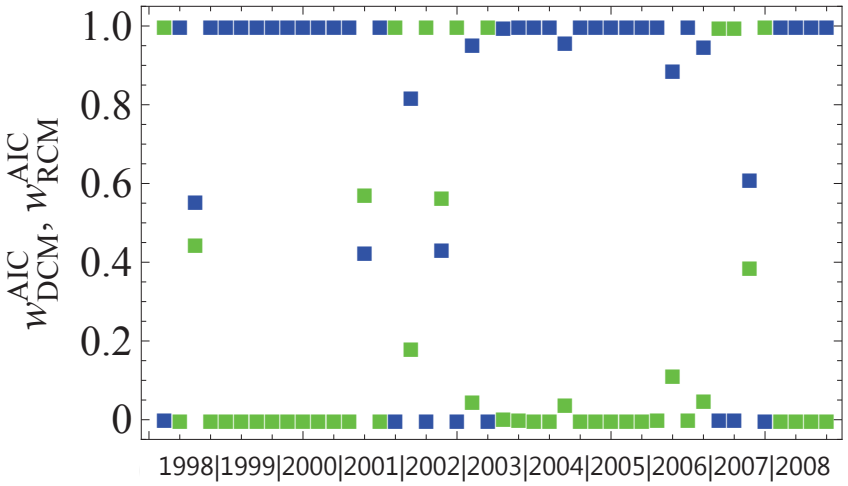

FIG. 14. Temporal evolution of Akaike weights for the DCM (blue) and the RCM (green).

$$AIC^t(\vec{\theta}^*) \equiv 2K^t - 2\ln \mathcal{L}^t(\vec{\theta}^*). \tag{41}$$

The model with the smallest value of AIC^t achieves the optimal balance between explanatory power (large log-likelihood) and parsimony (small numer of parameters) [12]. Across all snaphots, we found that the DRG is never the best model, while the DCM and RCM sometimes compete. To better discriminate between two competing models, the *Akaike weights* can be used. Given two models 1 and 2, we can calculate the two values (for each time period)

$$\Delta_i^t \equiv AIC_i^t - \min\{AIC_1^t,\ AIC_2^t\},\ i = 1,\ 2 \tag{42}$$

to define the Akaike weights [12] as

$$w_1^t \equiv \frac{e^{\frac{-\Delta_1^t}{2}}}{e^{\frac{-\Delta_1^t}{2}} + e^{\frac{-\Delta_2^t}{2}}},\ w_2^t \equiv \frac{e^{\frac{-\Delta_2^t}{2}}}{e^{\frac{-\Delta_1^t}{2}} + e^{\frac{-\Delta_2^t}{2}}}. \tag{43}$$

By definition, $w_1 + w_2 = 1$. A value $w_i \approx 1$ implies that model i strongly outperforms the other model, which can therefore be discarded. Values $w_1 \approx w_2 \approx 1/2$ imply that both models are very similar, and neither can easily be discarded.

Our results are shown in fig. 14 for all shapshots. For four snapshots the two models compete (having the weights near the central value of 0.5), so that they should both be retained and some more refined form of multimodel inference would be needed (the so-called *multimodel average* [12]). However, for most of the remaining snapshots the DCM outperforms the RCM, which is therefore less effective in explaining the observed topology. Given the extreme values of the Akaike weights, AIC seems to classify RCM as an overfitting model most of the time, pointing out that the correct amount of information to use is encoded in the degree sequences only and that all the remaining higher-order structures (dyadic and triadic structure) should be considered as non trivial patterns revealing the self-organization of the DIN.

So, even if the topological information introduced by adding the local reciprocity structure to the contraints would seem not to be excessive, most of the time AIC classifies it as redundant and suggests that the optimal level of description is the one achieved by controlling for the in- and out-degree sequences, confirming that all higher-order patters starting from the dyadic ones should be regarded as significant. Clearly, in order to filter out the dyadic effect from the triadic abundances, and select only the triads which cannot be explained in terms of a combination of dyads, the RCM remains the model to use.

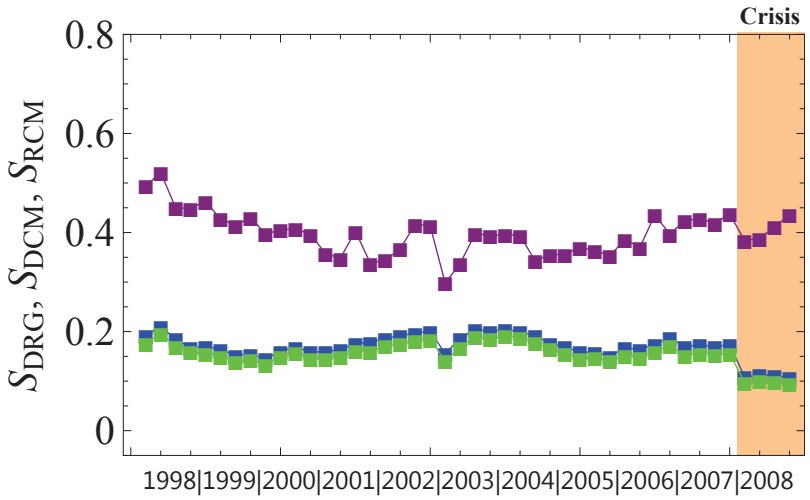

FIG. 15. Entropy for the three null models DRG (purple), DCM (blue) and RCM (green).

VIII. ENTROPY

In order to measure how effective the chosen constraints are in 'narrowing' the ensemble of networks around the observed configuration, we calculated the Shannon entropy of the probability distribution induced by each null model [2, 3]. To obtain comparable results across the three null models, we normalized all entropies between 0 and 1, considering that the DRG and the DCM predict probability coefficients for the directed pairs while the RCM predicts probability coefficients for the dyads. This normalization results in the following definition:

$$S_{NM} \equiv -\frac{\sum_{A\in\mathcal{G}} P_{NM}(A)\log_2 P_{NM}(A)}{N(N-1)},\ NM = DRG,\ DCM,\ RCM. \tag{44}$$

The result is shown in fig. 15. The same observations made about the connectance are also valid for the DRG entropy: the magnitude of the change between t_{40} and t_{41} is of the same (or even lower) order as the previous ones, so that it would be difficult to detect the crisis in terms of an anomalous behavior of S_{DRG}. On the other hand, both the DCM and the RCM entropies show a clear (even if not dramatically large) jump between 2007 and 2008, somehow indicating the onset of the crisis. However, there is no clear indication, neither in the DRG prediction nor in the DCM prediction, of a pre-crisis period. This seems to indicate that the clues of the upcoming instability are detectable neither in the degree sequences, nor in the local reciprocity structure, but in higher-order statistics (triadic motifs and core-periphery division), once the dyadic structure is kept fixed (degree sequences). Also note the small difference between S_{DCM} and S_{RCM}. While in going from the DRG to the DCM there is clearly a large information gain, there is only a much smaller gain in going from the DCM to the RCM. This explains why AIC most of the time indicates that the information gained by the RCM over the DCM is not enough to justify the introduction of the additional model parameters.

ACKNOWLEDGEMENTS

D. G. acknowledges support from the Dutch Econophysics Foundation (Stichting Econophysics, Leiden, the Netherlands) with funds from beneficiaries of Duyfken Trading Knowledge BV, Amsterdam, the Netherlands.

[1] van Lelyveld, I. & in't Veld, D. Finding the core: network structure in interbank markets, *DNB Working Paper* **348** (2012).

[2] Shannon, C. A Mathematical Theory of Communication, *Bell System Tech. Jour.* **27**, 379423, 623656 (1948).
[3] Jaynes, E. T. Information Theory and Statistical Mechanics, *Phys. Rev.* **106**(4) (1957).
[4] Holland, P. & Leinhardt, S. *Sociological Methodology*, Heise (ed.) (San Francisco, 1975).
[5] Wasserman, S. & Faust, K. *Social Network Analysis* (Cambridge University Press, New York, 1994).
[6] Park, J. & Newman, M. E. J. The statistical mechanics of networks, *Phys. Rev. E* **70**, 066117 (2004).
[7] Garlaschelli, D. & Loffredo, M. I. Maximum likelihood: extracting unbiased information from complex networks, *Phys. Rev. E* **78**, 015101(R) (2008).
[8] Squartini, T. & Garlaschelli, D. Analytical maximum-likelihood method to detect patterns in real networks, *New. J. Phys.* **13**, 083001 (2011).
[9] Garlaschelli, D. & Loffredo, M. I. Patterns of link reciprocity in directed networks, *Phys. Rev. Lett.* **93**, 268701 (2004).
[10] Garlaschelli, D. & Loffredo, M. I. Multispecies grand-canonical models for networks with reciprocity, *Phys. Rev. E* **73**, 015101(R) (2006).
[11] Newman, M. E. J., Forrest, S. & Balthrop, J. Email networks and the spread of computer viruses, *Phys. Rev. E* **66**, 035101(R) (2002).
[12] Burnham, K. P. & Anderson, D. R. *Model Selection and Inference: A Practical Information-Theoretical Approach* (Springer-Verlag, New York, 1998).
[13] Picciolo, F., Squartini, T., Ruzzenenti, F., Basosi, R. & Garlaschelli, D. The role of distances in the World Trade Web, in *Proceedings of the 2012 IEEE/ACIS 11th International Conference on Computer and Information Science*, 784-792 (IEEE, 2013).